Decoding Collagen Triple Helix Stability by Means of Hybrid DFT Simulations

Decoding Collagen Triple Helix Stability by Means of Hybrid DFT Simulations

Journal:	*The Journal of Physical Chemistry*
Manuscript ID	jp-2019-052228.R1
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Cutini, Michele; Universita degli Studi di Torino, Dipartimento di Chimica Bocus, Massimo; Ghent University, Center for Molecular Modeling; University of Turin, Department of Chemistry Ugliengo, Piero; Universita degli Studi di Torino, Dipartimento di Chimica

Decoding Collagen Triple Helix Stability by Means of Hybrid DFT Simulations

Michele Cutini,[1]* Massimo Bocus[1,2], and Piero Ugliengo[1]*

[1]*University of Torino, Department of Chemistry and NIS (Nanostructured Interfaces and Surfaces) Center, Via P. Giuria 7, 10125 Turin – ITALY*

[2]*Present address: University of Ghent, Center for Molecular Modeling, Technologiepark 46, 9052 Zwijnaarde, Belgium*

* *e-mail: michele.cutini@unito.it; piero.ugliengo@unito.it*

Abstract

Collagen is a protein family defined by a triple helix motif, which comprises roughly one-third of the total human protein content. Decoding the reasons underlying the stability of the collagen triple helix is of both fundamental and applicative relevance, for instance, to guide collagen protein engineering. In principle, full quantum mechanical approaches based on density functional theory (DFT) are ideal to study the subtle physico-chemical features of collagen. Unfortunately, the huge size of the protein prevents the straightforward application of DFT to realistic collagen protein models. In this paper, we propose a new realistic model of the collagen protein based on a periodic approach. The protein model exploits the intrinsic symmetry of the collagen triple helix, dramatically lowering the cost of the simulations. This allows using accurate hybrid DFT simulations (B3LYP-D/TZP) for systematic studies of the collagen protein features. We have tested the proposed model/level-of-theory combination to analyze the well-known proline-conformation/collagen-stability relationship. For this purpose, we have performed an extensive conformational analysis of proline ring within the protein, clarifying some of the reasons linking specific ring conformations to helix positions. Throughout our data analysis, we have also obtained "for free" the collagen inter-strand binding energy. Simulation results demonstrate that London dispersion interactions play the dominant role in the whole helix stability. The good agreement with the experimental data validates the use of the proposed model/level-of-theory to assist the active field of collagen-like peptide synthesis.

Introduction

Collagen (COL) is a protein family, which includes 28 members to date. COLs are made by three identical (homo-trimeric) or different (hetero-trimeric) polypeptide chains.[1,2] These chains are wrapped together giving rise to the well-known COL triple helix, see Figure 1A.

The various proteins of the family are labelled with roman numerals (I, II, etc.) in the chronological order of their discovery and grouped in categories mainly based on: (i) their propensity of associating in well-organized supra-molecular assemblies, e.g. COL fibrils; (ii) the number of triple helical domain interruptions.[2] The fibril-forming COLs present mainly the triple helical domain with only two shorts non-helical regions at the ends, called telo-peptides. The ~90% of human COL is fibril-forming COL (collagen type I, II, III, V and XI) and comprises roughly one-third of the total human protein content. These are found abundantly in vertebrate tendons, bones, skin, cartilage, cornea, blood vessels, etc.[2]

As mentioned before, the COL triple helix motif defines the entire family of collagens. A seminal work on COL structure determination was first proposed by Ramachandran and Kartha in 1954,[3] and then refined in the later years[4] by contributions from Rich and Crick[5] and Cowan et al.[6] In each individual COL protein, usually called tropocollagen, the helical domain is composed by three parallel polypeptides strands wrapped together, see Figure 1A. Each single strand (ss-COL), which is coiled to allow the protein aggregation, see Figure 1B, has also an inner helicity. In details, this is a poly–proline type II (PPII) helicity, see Figure 1C.[7–9] For each strand, the primary structure follows a strict repetitive pattern of triplets where Glycine (Gly) occurs always in the same position (Gly-X-Y). The aminoacids in X and Y positions are exposed to solvent while Gly is buried inside the helix core, Figure 2A. The three ss-COLs wrap together with one residue staggering to close packing the structure, Figure 2B. This allows the establishing of H-bonds between the carboxylic group of the aminoacids in the second position (X) and the N-H of Gly, Figure 2C, 2D.[2]

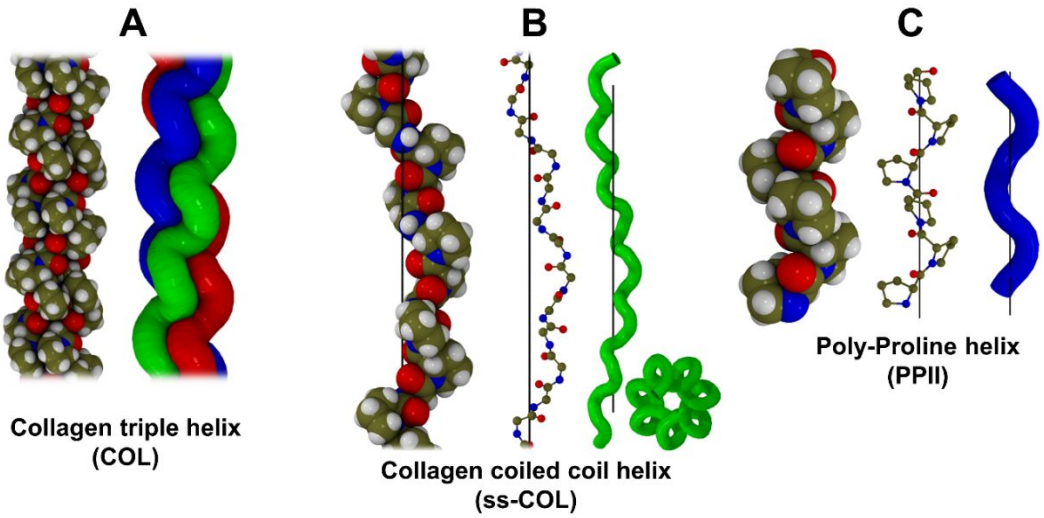


Figure 1. **A**: COL triple helix as van der Waals radii (vdW) and with single strands represented in different colours. **B**: Single polypeptide chain (ss-COL) as it is within the COL helix, vdW, ball and stick (backbone atoms only), and tube (side and top views) representations. **C**: Single polypeptide chain (all proline) in PPII conformation in a vdW, ball and stick, and tube representation. Colour code: Carbon/cyan, Oxygen/red, Nitrogen/blue and Hydrogen/light-grey.

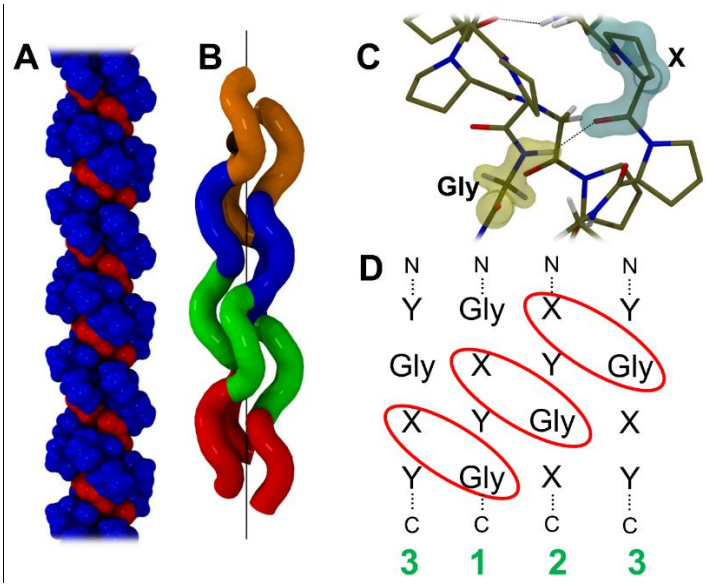


Figure 2. **A**: COL side view with Glys in red and X and Y amino acids in blue. **B**: COL side view (tube) with each Gly-X-Y triplet coloured differently to highlight the staggering between the chains. **C-D**: Details on the inter-chain H-bonds. C: The residue involved in the H bonds are highlighted. D: Schematic representation of COL triple helix in which the residues involved in the H-bonds are circled in red. The chains are numbered in green, e.g. 1, 2 and 3, and their direction is indicated, with C or N, indicating the carboxylic and amino termini of the protein, respectively. The strand number 3 is repeated two times for clarity on the H-bond pattern.

Proline (Pro) and (2S,4R)-4-Hydroxyl-proline (Hyp) are the most common aminoacids in the second and the third position (28% and 38%, respectively). In all collagens, the Gly-Pro-Hyp (GPO) triplet is the most common triplet (10.5%),[10] and Pro and Hyp represent the ~22% of all residues. The high content of Pro and derivatives is thought to give stability to the triple helix. This type of stabilization is known to be manly entropic. Each Pro-rich strand is pre-organized to a PPII helix in the unfolded state, hence to a geometry more similar to the folded triple helix than a random conformation.[1,2]

Also the Pro and derivatives side-chain conformation plays an important role in the stability of COL.[10] Pro is known to have two stable side chain conformations, hereafter indicated as DOWN (D) and UP (U). For *trans* Pros in gas phase,[11,12] flipping from one pucker to another costs less than 3 kJ·mol^{-1}, in favour of the D state, thus both states are almost equally populated at room temperature, as it is in proteins.[13] Interestingly, electron-withdraw substituents on certain position of the pyrrolidine ring can alter the Pro puckering propensity.[14] Hyp, envisaging the OH group, is a natural example substitution on the pyrrolidine ring. In this case, the OH substitution takes place on a ring position which leads to the stabilization of the U puckering. The enzymatic hydroxylation of the Pros in Y position enhances the stability of the COL triple helix.[1] A similar effect is found for Pros in X position. Indeed, synthetic COL-like peptides, in which the Pros in X are substituted to enhance the stability of the D puckering, are more stable than the non-substitute COL.[1,2]

A possible explanation for these phenomena derives from the strong correlation between the main ring dihedral angle of Pro and derivatives as well as its side chain conformation. It seems that Pro and derivatives with U/D puckering in the Y/X positions pre-organize the main chain dihedral angles to fit into a COL triple helix geometry without too much deformation and entropic cost.[1,13,15] The opposite is true as well, i.e. Pro derivatives on Y/X position with enhanced tendency to D/U puckering destabilizes the triple helix. This simple statements are known as the *propensity based hypothesis*. Following this hypothesis, it becomes possible synthetizing several hyperstable collagen-like peptides,[1,2] even if that did not work in all cases.[1,16–19]

In this paper we have analyzed the phenomena guiding specific side chain conformations of Pros buried in collagen before and after the Pro hydroxylation process. To achieve this goal, we have investigated the effect of Pro ring conformations on the stability of realistic collagen triple helical polymers by means of molecular simulations. The vast majority of modeling works on collagen features adopted, as theoretical framework, the molecular mechanics paradigm.[20–25] In this work, we instead choose a quantum mechanical approach based on density functional paradigm. This class of methods is parameters free and solve the Schrödinger equation through the ground state electron density which, through a proper functional, determines the system ground state energy. Among

functionals, the hybrid ones provide better results due to the inclusion, in their definition, of a fraction of exact Hartree-Fock exchange. In the present work, we adopt the B3LYP[26–28] associated to the Grimme's D3^{ABC} correction for including London dispersion interactions,[29,30] not accounted for by the plain functional. The B3LYP-D3^{ABC} level associated with Gaussian basis set of polarized valence triple-zeta quality VTZP ensure enough accuracy of the calculations,[31] crucial to detect small energy differences for the Pro side chain conformers. We have considered homo-trimeric COL polymers made by the repetition of Gly-Pro-Hyp (GPO) and Gly-Pro-Pro (GPP) aminoacidic triplets. We have focused on the role of the side chain of Pro and derivatives in altering: i) the stability of the triple helix, ii) the interaction energy between strands. To have more insights for our analysis we have also evaluated the preferential state of Pro when buried in different protein structures, e.g. in a single Pro residue, in stretched polymer (Figure 1C), as a coiled coil structure (Figure 1B), and in collagen itself (Figure 1A).

The results obtained are in very good agreement with the available experimental data, indicating the reliability of the chosen protein-model and level-of-theory combination.

Molecular Models

Collagen Triple Helix (COL and ss-COL)

Collagen exhibits different type of helicities depending on the amino acid composition.[1,2] Currently, there is agreement that for high content of Pros and derivatives, collagen has a 7/2 helicity. This means that 7 aminoacid triplets fit into two helix turns, see Figure 3A. The usual translation repeat of such helix is ~20 Å. Pro-free zone of collagen seems to follow a looser helicity, e.g. 10/3. This has 10 aminoacid triplets in three helix turns, with a ~28 Å long unit cell. In this paper, due to the high Pro content of the chosen collagens compositions, we considered only collagens with 7/2 helicity and we modelled them starting from two assumptions: i) *the collagen protein is an infinite polymer (1D periodic)*; ii) *we impose roto-translational symmetry to the 1D collagen model*. As for point i), we believe, the periodic model is more representative of the real protein than any finite and short triple helical molecule usually adopted when studying collagen. The infinite 1D periodic model also avoids the need for capping substituents (to avoid inter-chains interaction of free COOH and NH_2 groups at the molecule edges) at the peptide termini, which may introduce artefacts in the results. Point ii) allows to define the polymeric 1D collagen model by one aminoacidic triplet that is replicated by the symmetry operators giving the whole triple helix structure (see Figure 3A, 3B, 3C, 3D and Figure 3E for a detailed description).

It is worth noting that exploiting symmetry constraints within the collagen 1D model, speeds up the simulation notably. The expected speed-up factor depends on the type of helicity (number of symmetry operators) imposed to the collagen model. For instance, for a 7/2 or 10/3 helix the speed up factors are 7 or 10 with respect to a plain P1 symmetry polymer. A further speed-up factor is on the number of steps needed to converge the geometry optimization. Indeed, symmetry purifies the potential energy surface of the system, reducing the number of steps necessary to reach stationary points. Even more important is the cost reduction when simulating the vibrational spectrum (not shown here) in which only the irreducible symmetry elements of the Hessian matrix are evaluated.

We also took into consideration the geometry of a collagen single strand as it is within the COL triple helix, e.g. ss-COL. A helix of this type is a 7/1 helix with a ~60 Å long unit cell see Figure 1B (axial translation per triplet of 8.6 Å with 7.0 residues per turn). This 7/1 helix is graphically reported as a coloured line in Figure 3A.

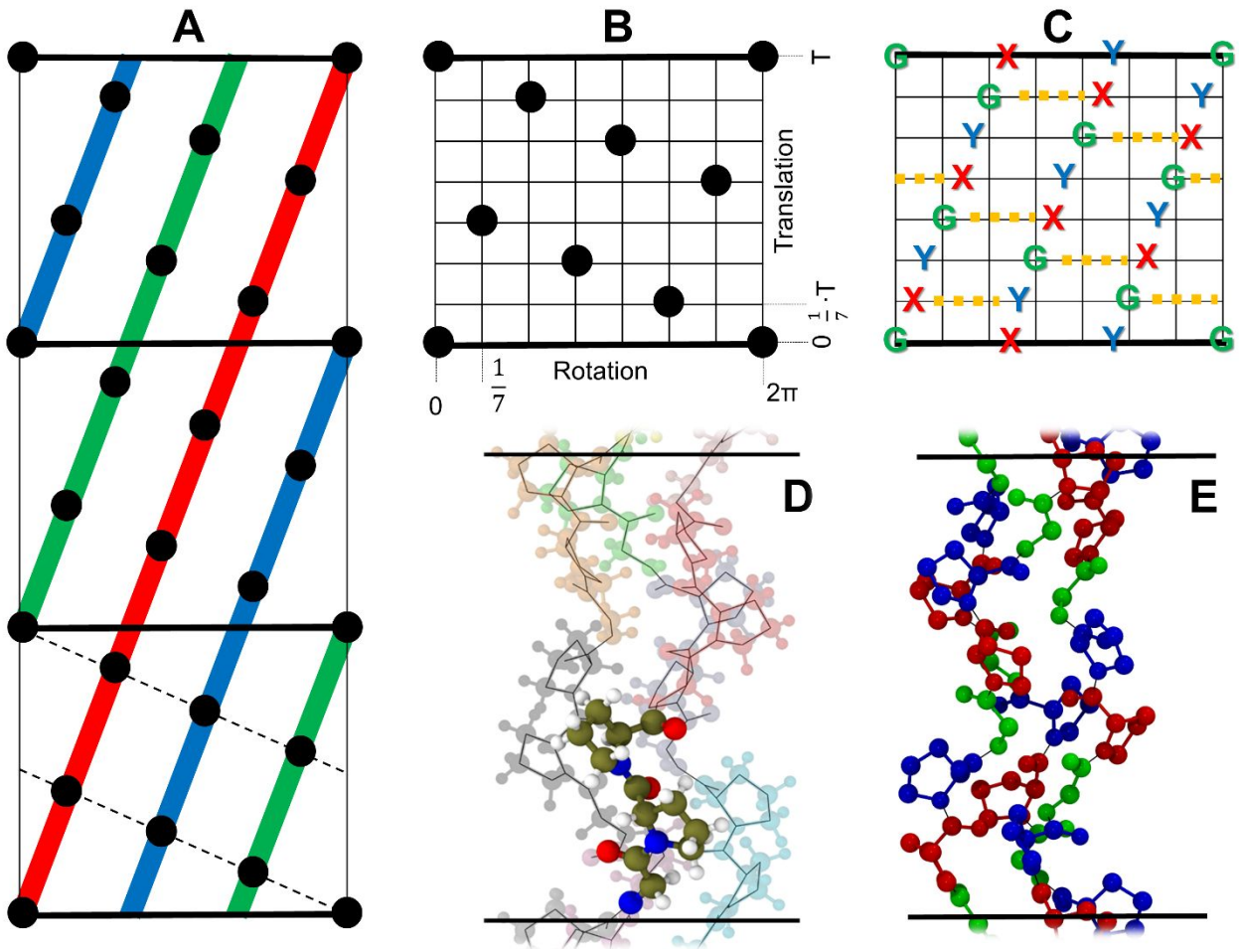


Figure 3. **A**: Radial projection of the 7/2 helix (7 scattering units within two helical turns), e.g. COL. Black circles are the Gly–X–Y triplets. We indicated the peptide chain direction using colored solid lines. Each independent colored solid line is a 7/1 helix, e.g. ss-COL. Dotted black lines indicate the helical rotation. **B**: Radial projection of the built using the HELIX keyword as implemented in the CRYSTAL17 code. Black circles are the Gly–X–Y triplets. **C**: Same as B with the aminoacids explicitly reported. The inter-strand H bond are reported as yellow dotted lines. **D**: Lateral view of helix represented in B with GPP composition. The irreducible by symmetry atoms within the polymer unit cell are in full colors. **E**: Lateral view of helix represented in C with GPP composition.

Poly-Proline Type II Helix (PPII)

We have considered the poly-proline type II helical geometry (PPII) as a model for a stretched single collagen protein strand. The PPII helix is a left handed 3/1 helix with a ~9 Å long unit cell, see Figure 1C (axial translation per triplet of 3.20 Å with 3.0 residues per turn). In this helix all peptide bonds are in *trans* configuration with ideal dihedral angles (ϕ, ψ, ω) of (-75°, 145°, 180°). This system has been recently investigated with DFT simulations by us, as a single polymer,[32] and in interaction

with hydroxyl-apatite surface,[33] and, by others, with the same methodology adopted here, in its crystal form.[34]

Molecular Models (MOL)

To assess the puckering propensity of the Pro and Hyp aminoacids in a random coil protein geometry, e.g. without geometrical constrains as imposed in COL, ss-COL and PPII cases, we considered also the following molecular models (MOL): i) Proline (Pro), ii) (2S,4R)-4-Hydroxyl-Proline (Hyp), iii) (2S,4R)-4-Fluoro-Proline (Flp), and iv) N-Methyl-(2S,4R)-4-Fluoro-Proline (M-Flp). In all molecules, the carboxyl group is in the *trans* conformation, the OH group of the carboxyl group makes an intramolecular H-bond with the nitrogen lone pair. This conformation has been demonstrated to be the most stable one in the case of Pro in gas phase,[12] and it was extensively studied by us in Ref [32].

Conformational Analysis

In this work we have investigated homo-trimeric COLs made by the repetition of an identical aminoacidic triplet G-X-Y (Figure 4A). We considered two different triplets, e.g. Glycine-Proline-Proline (GPP) and Glycine-Proline-(2S,4R)-4-Hydroxyl-Proline (GPO). We also investigated the same compositions as single independent strands with a PPII geometry. We explore all the conformational variability of the models: the main dihedral angles of the polypeptide chains defining the COL/PPII helix, are locked by the helical symmetry in the COL case and by translational symmetry for the PPII case; thus, the conformational freedom of our models arise from the side chain flexibility only. Regarding the MOL cases, several degrees of freedom are available, but for our purposes, we focused only on the side chain conformations.

Glycine (Gly), with a hydrogen atom as side chain, does not possess any conformational variability. Proline (Pro) and (2S,4R)-4-Hydroxyl-Proline (Hyp) side-chains are well-known to have two stable conformations, e.g. UP(U) and DOWN(D). These have been extensively studied with NMR spectroscopy[35–39] and, theoretically, in an all-Pro PPII polymer within our group.[32] Those conformations are called U/D, in association with positive/negative value of the χ^2 dihedral angle (see Figure 4B). The U/D conformer has the C_γ on the opposite/same side of the carbonyl group with respect to the pseudo-plane defined by the other ring atoms (N, C_α, C_β and C_δ), see Figure 4B and 4C. Electron withdraw substituents at Cγ on R/S position (see Figure 4D) of pyrrolidine ring stabilizes U/D puckering in Pros. OH group in Hyp is in R position, thus Hyp stable side chain conformation is U. The presence of a hydroxyl group on the Hyp side chain ring adds a further degree of freedom to our system. Three stable conformers exist associated with the staggered conformations of the

hydroxyl hydrogen with respect to the substituents on the adjacent carbon. We will call those conformations A, B and C and the dihedral angle H-C_γ-O-H will be indicated as ξ (see Figure 4E).

In our COL/ss-COL/PPII models, the asymmetric atoms within the polymer unit cell (those not replicated by symmetry operators within the polymer 1D cell) are those belonging to one aminoacidic triplet. The two pyrrolidine rings within one aminoacid triplet can be D or U and the models will be named as AB, where A/B is the puckering of the pyrrolidine ring in the X/Y position. In the case of GPO, a further letter is added to the name identifying the orientation of the OH group. Therefore, for GPP, the possible conformers are four, e.g. DD, DU, UD and UU. For GPO they vary depending on the system (COL/ss-COL or PPII). In the case of COL/ss-COL, they are twelve, the four puckering conformers DD, DU, UD and UU for each of the three orientations of the OH group (A, B and C), e.g. DDA, DUA, UDA, UUA, DDB, DUB, UDB, UUB, DDC, DUC, UDC and UUC. In the case of PPII, an intra-chain H-bond forces the orientations of the OH group. It leads to the exploration of also eclipsed orientations, named in order as D, E and F, see Figure 4E.

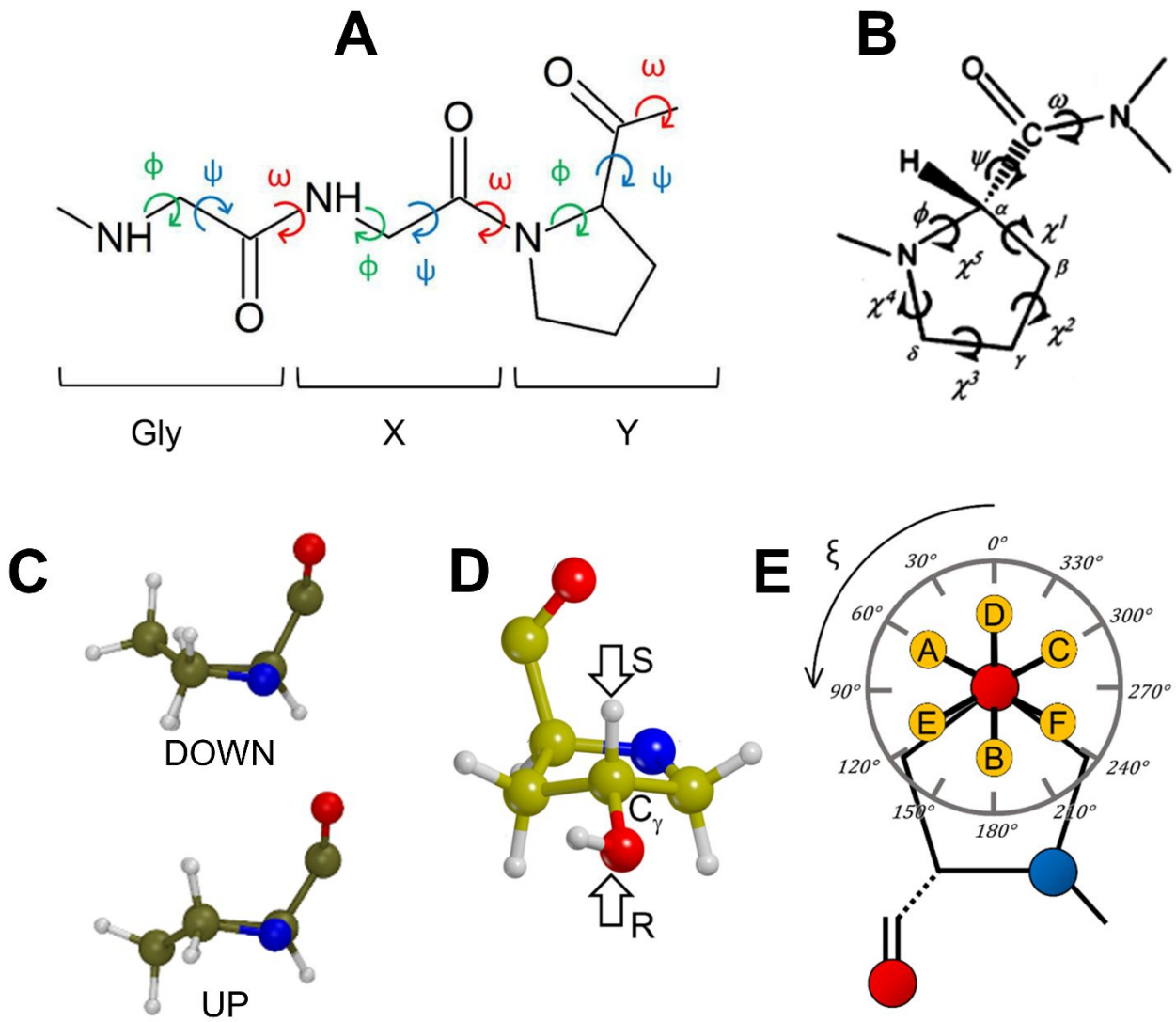


Figure 4. **A**: Definition of the dihedral angles within the COL/ss-COL/PPII triplet. **B**: Definition of the atoms and the dihedral angles within the Pro residue. **C**: Ball-and-stick images of Pro as extracted from the COL/ss-COL/PPII geometry in the DOWN and UP conformation. **D**: Ball-and-stick images of Hyp as extracted from the COL/ss-COL/PPII geometry. The S and R position on the Cγ atom are highlighted. **E**: Schematic representation of Hyp as seen along the C_γ-O bond. Colour code: Covalent bonds are in black lines, oxygen in red, nitrogen in blue, hydroxyl hydrogen in yellow. The nomenclature of the conformations of the hydroxyl hydrogen are shown, together with the values of the H-C-O-H dihedral angle (ξ).

Computation Details

All DFT calculations in the present work were performed with a development version of the CRYSTAL17 code.[40] We employed the B3LYP hybrid functional,[26–28] coupled with the Ahlrichs' polarized valence triple-zeta (VTZP) basis set for all the atoms,[41] with the exception of hydrogen. For the latter, we used a 3-1G(p) basis set.[42] For Fluorine atom we employed a 7-311G basis set. We introduced the semi-classical D3 dispersive correction as well,[43] including the Axilrod–Teller–Muto (ATM) three bodies term ($D3^{ABC}$).[29,30]

In the geometry optimizations, the atomic positions and cell vectors were relaxed using the analytical gradient method. The Hessian was upgraded with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm.[44–46] The tolerances for the convergence of the maximum allowed gradient and the maximum atomic displacement were set to the default values. The Γ-centered k-point grid was generated via the Monkhorst−Pack scheme[47] with a system dependent number of k points. The shrink factor for all polymer calculations was set to 4. To help convergence of the SCF, the Fock/KS matrix at a cycle was mixed with 30% of the one of the previous cycle.[48] The electron density and its gradient were integrated over a pruned grid consisting of 974 angular and 75 radial points generated through the Gauss–Legendre quadrature and Lebedev schemes.[49] Tolerances were set to 10^{-6} for Coulomb overlap, Coulomb penetration, exchange overlap and exchange pseudo-overlap in the direct space, and 10^{-14} for exchange pseudo overlap in the reciprocal space.

The graphical visualization and structural manipulation of structures was performed with MOLDRAW version 2.0.[50] Images were rendered with VMD version 1.9.3.[51]

Definition of the Computed Quantities

The binding energy between COL strands (BE^*) is defined as:

$$BE^* = E(\text{ss-COL//COL}) - E(\text{COL//COL}) \quad (1)$$

where the symbol following the double slash identifies the optimized geometry at which the energy has been computed. Therefore, E(ss-COL//COL) is the energy of a single strand (ss-COL) extracted from the COL triple helix (COL), in the COL fully optimized geometry; in other words, the geometry of the single strand ss-COL is retained in that resulting for the triple helix COL. E(COL//COL) is the energy of the COL polymer in its fully optimized geometry. BE^* is defined as a positive quantity for a bounded system.

We point out that in equation 1), there is no need of the multiplicative factor 3 for the single strand E(ss-COL//COL) energy, as one might expect. This is because a single strand collagen has a different helicity and unit cell than a triple helical collagen, e.g. 7/1 and 7/3, respectively. This leads the unit cell of the single strand collagen to be made by seven aminoacid triplets as a triple helical collagen. In the former case, as already mentioned in the *molecular models* section, the unit cell length is three times longer than the latter case.

We corrected the BE^* for the BSSE, e.g. BE^{*C}, as following:

$$BSSE = E(ss\text{-}COL//COL) - E(ss\text{-}COL[ss\text{-}COL]_2//COL) \quad (2)$$

$$BE^{*C} = BE^* - BSSE \quad (3)$$

In which the term $E(ss\text{-}COL[ss\text{-}COL]_2//COL)$ indicates the energy of a single collagen strand surrounded by two strands of ghost atoms, in which the nuclei are substituted by their basis set function.

The energy differences between conformers (for instance A and B) in a ss-COL and COL geometry are interconnected by the BE^* and BSSE terms as following:

$$\Delta E(COL) = E(COL_A//COL_A) - E(COL_B//COL_B) \quad (4)$$

$$\Delta E(ss\text{-}COL) = E(ss\text{-}COL_A//COL_A) - E(ss\text{-}COL_B//COL_B) \quad (5)$$

$$\Delta BE^{*C} = BE^{*C}_A - BE^{*C}_B \quad (6)$$

$$\Delta BSSE = BSSE_A - BSSE_B \quad (7)$$

$$\Delta E(COL) = \Delta E(ss\text{-}COL) - \Delta BE^{*C} - \Delta BSSE \quad (8)$$

Clearly the energy differences between conformers (A and B) in the MOL and PPII geometries are computed as:

$$\Delta E(MOL) = E(MOL_A//MOL_A) - E(MOL_B//MOL_B) \quad (9)$$

$$\Delta E(PPII) = E(PPII_A//PPII_A) - E(PPII_B//PPII_B) \quad (10)$$

The contribution to the BE* values due to the London dispersion interactions is easily computed by the $D3^{ABC}$ separate contribution to the total B3LYP-$D3^{ABC}$ energy.

Results and Discussion

Molecular gas-phase models (MOL)

We have investigated the puckering propensity of Pro, Hyp, Flp and M-Flp molecules. As expected by previous studies,[12] the D pucker is the most stable one for Pro, e.g. 2.6 kJ·mol^{-1} more stable than U. Conversely, the U puckering is the most favored for Hyp and Flp molecules. In these cases, the U→D process costs 7.7 and 13.2 kJ·mol^{-1}, respectively. Flp prefers the U conformation as the fluorine atom is intra-H-bonded with the N-H group, which is absent in the D conformation, thus stabilizing the U conformer. Substituting N-H with N-CH_3 reduces the puckering cost of flipping U→D to 4.6 kJ·mol^{-1}, closer to Hyp case.

Gly-Pro-Pro (GPP) gas-phase models

The structures (cell length and internal coordinates) of the four stable Gly-Pro-Pro (GPP) unfolded, PPII, and folded, COL, collagen protein conformers were optimized at the B3LYP-D3^{ABC} /VTZP level. The results and the analysis are reported in Figure 5-6, Table 1 and Table S1 of the supplementary information. The results for the GPP compositions in the PPII polymer geometry are reported along with those of the all-Pro PPII polymer, in Table 1. The all-Pro PPII polymer was extensively studied in a previous paper published by some of us[32], with an equivalent level of theory. The COL inter-strand interaction energy (BE^{*C}) and the energy difference between single COL strand as deformed in the COL geometry, e.g. ΔE(ss-COL) (see equation 5), are reported in Table 2.

The PPII case

In the PPII single chain polymer geometry, the most stable conformation envisages all Pros in a DOWN puckering one (conformer DD, see Table 1). From DD, the energy cost to side chain flipping from DOWN to UP (D→U) depends on the position of the flipped Pro (X or Y position, DD→UD (7.6 kJ·mol^{-1} triplet^{-1}) and DD→DU (5.4 kJ·mol^{-1} triplet^{-1}) process, respectively). This indicates that the presence of Gly introduce a different propensity on Pro pucker within the PPII polymer: the U conformation is more accessible when Pros occupies the Y position. This can be explained by the geometrical distortion induced by Gly. For an all-Pros PPII polymer in the DDD state, all Pros have ϕ is -72.5° and ψ is 157.2° (see Table 1). Due to the Gly presence ϕ splits in two sets, namely -79.3° and -68.4° for Pro in X and Y, respectively. The ϕ value expected for a U Pro is ~ −58.7°,[13] thus the ϕ value of Pro in Y is closer to the expected one than for Pro in X position.

The COL case.

Moving from a single strand PPII to a COL polymer, we predicted changes in the relative stability ΔE(COL) of the conformers. Interestingly, the most stable conformation for GPP in the COL polymer is DU (see second row of Table 2). This result is in good agreement with experiments: from X-ray structures on triple helical GPP collagen-like peptides, Pros in Y are, for most of the cases, in the U state (D:U = 0:10,[52] or 1:6[53]) and Pros in (X) are 100% of the times in the D state (D:U ration = 10:0,[52] or 7:0[53]).[52–54]

Also, the geometry of the DU COL agrees well with the experiments. The average unsigned deviation of the backbone dihedral angles is only 4°, see Table S1. Moreover, also the inter-strand contact length (Hb in Table S1) is well reproduced by the simulations, giving slightly shorter contacts than the experimental ones, probably due to BSSE and/or the thermal expansion, both neglected by our approach. When considering the ss-COL, it turned out that the U conformation occupying the Y position brings about only 1.3 kJ·mol^{-1} triplet^{-1} of instability for DU compared to the DD most stable state (see ΔE(ss-COL) in Table 2). Interestingly, the DU state becomes the most stable one when ss-COL assembles in the COL polymer. This extra stabilization is due to the inter-strand energy (BE*^{C}, see Table 2) which is slightly higher for the DU state than for the other cases. Remarkably, this extra stabilization is mainly due to London dispersion components of the whole interaction energy.

To understand the effect of the collagen-structure/Pro-side-chain relation, we have investigated the Pro conformers in different protein organizations as mimicking a hypothetical folding process of collagen. This would start with the protein in a random coil conformation (step I). In our simulated process the second step envisages the stretching of the random coil polymer into a linear filament (step II), which then deforms to allow the right conformation for aggregation (step III), finally realized with the other two chains to give the collagen triple helix (step IV). We cannot model step I, as the random coil spans a too large configurational space to be treated with our static methodology at full quantum mechanical level. Therefore, for step I, we simply refer to the free aminoacid constituent of the triplet presents in the extended polymers. In summary, the conformational state has been explored in each step, for:

- step I: a single aminoacid molecule (MOL)
- step II: a straight infinite polypeptide polymer (PPII)
- step III: a coiled coil single polypeptide chain (ss-COL)
- step IV: a triple helix (COL)

Figure 7A-B summarizes our results. Figure 7A shows the results for Pro when the remaining nearby Pro in the triplet is in the D state. We can state that: i) stretching a protein strand stabilizes the D

puckering in any triplet position, compared to a random protein conformation; ii) the Gly presence within the aminoacidic triplet stabilizes the Pro D/U state in X/Y, see comparison with all-Pro PPII polymer (Table 1); iii) coiling the straight polymer stabilizes the U pucker both in X and Y, as well as; iv) introducing the interaction between strands. Interestingly the effect of iii) and iv) is higher in the Y position than in the X one, so leading the Pro in Y to have a stable U pucker in COL.

Figure 7B shows the results for Pro when the remaining nearby Pro in the triplet is in the U state. The situation changes in this case, indeed: i) the stretched protein has similar pucker propensity than the random coil; ii) the Gly presence stabilizes the Pro D/U state in X/Y, as in the previous case; iii) the coiling of the strands and the interaction between them makes the D state less/more accessible in Y/X. Interestingly, the Gly presence, the coiling into a ss-COL geometry and the inter-strand interaction widen the stability gap between D and U states in X and Y positions, see Figure 7A and 7B. This indicate the clear propensity of the D pucker in X and vice versa the U pucker in Y.

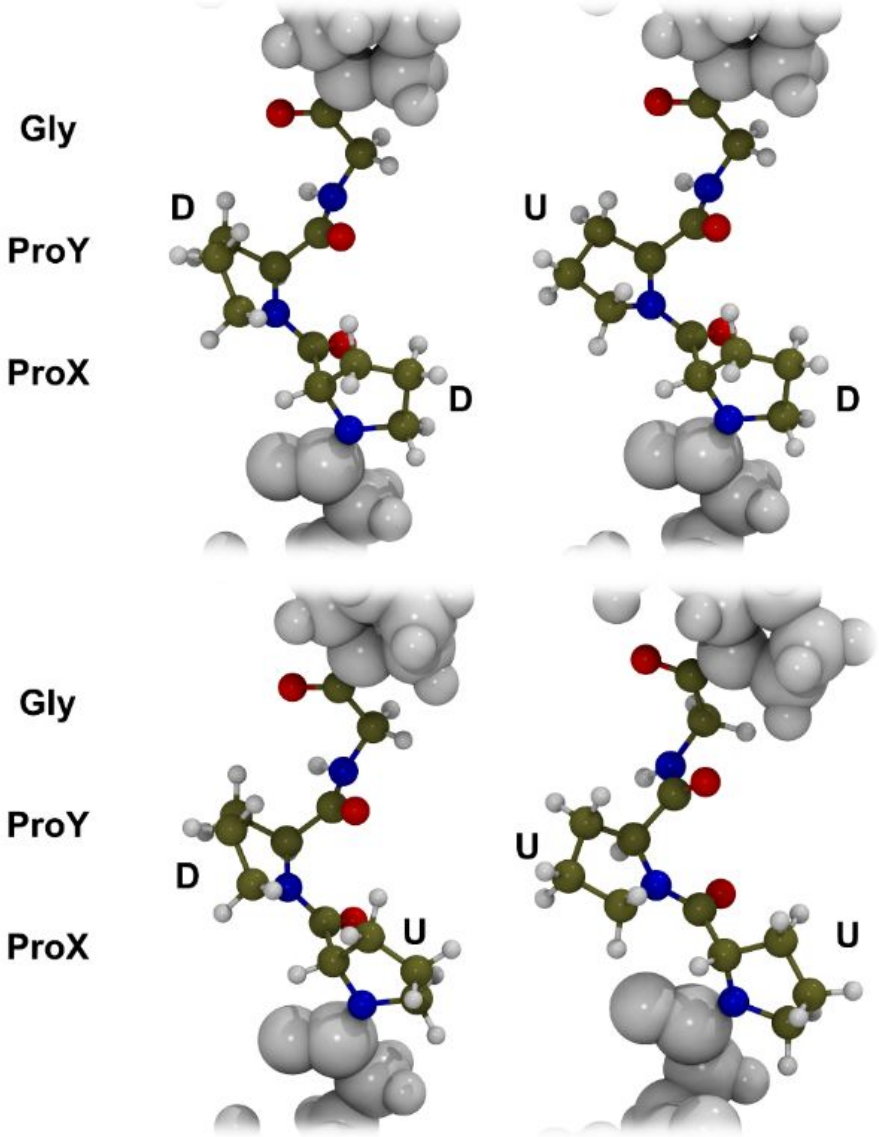


Figure 5. B3LYP-D3^{ABC}/VTZP optimized geometries of the GPP single strands in the PPII conformation. The amino acid asymmetric triplet is highlighted in ball-and-stick representation. The letters close to the Pro rings refer to the puckering of its side chain.

Table 1. Geometrical parameters and energies of the GPP and all-Pro aminoacids in a PPII single strand geometry. Angles in degrees, cell length *a* in Å, relative energies with respect the most stable state in kJ·mol^{-1}·triplet^{-1}.

	Gly (or Pro)				Pro (X)				Pro (Y)					
name	φ	ψ	ω	χ^2	φ	ψ	ω	χ^2	φ	ψ	ω	χ^2	*a*	ΔE(PPII)
Gly-Pro-Pro (Figure 5)														
DD	**-83.3**	**163.1**	**173.7**		**-79.3**	**161.2**	**162.9**	**-35.5**	**-68.4**	**161.3**	**159.8**	**-35.9**	**9.33**	**0.0**
DU	-84.1	162.9	175.1		-78.2	156.1	160.2	-35.4	-61.8	158.5	160.4	31.5	9.27	5.4
UD	-81.3	159.6	170.6		-75.4	162.2	163.2	21.3	-68.7	162.4	159.6	-35.8	9.24	7.6
UU	-63.9	131.2	173.4		-59.0	130.9	175.7	38.5	-58.8	144.0	175.7	38.2	8.75	8.5
Pro-Pro-Pro[32]														
DDD	-72.5	157.2	166.6	-35.1	-72.5	157.2	166.6	-35.1	-72.5	157.2	166.6	-35.1	9.12	0.0
UDD	-60.2	146.7	176.0	36.9	-73.8	154.6	173.8	-36.7	-70.0	138.1	167.1	-35.4	8.96	5.0
UUD	-57.3	131.3	174.2	38.1	-61.5	148.3	179.0	37.8	-69.6	137.9	170.3	-36.1	8.77	7.0
UUU	-57.5	133.3	174.5	38.5	-57.5	133.3	174.5	38.5	-57.5	133.3	174.5	38.5	8.61	7.4

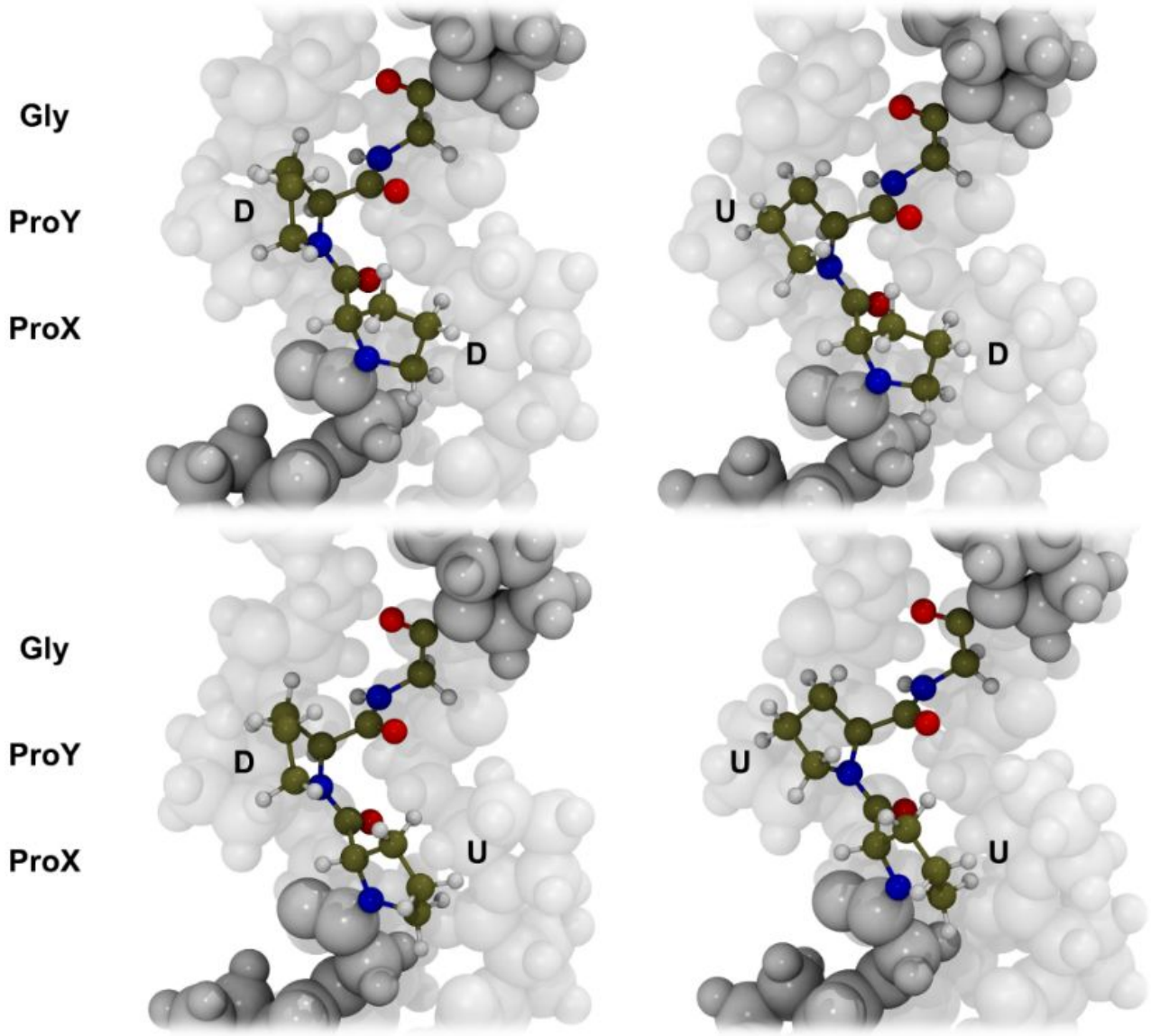


Figure 6. B3LYP-D3^{ABC}/VTZP optimized geometries of the GPP COL models. The amino acid asymmetric triplet is highlighted in ball-and-stick representation, whilst the other atoms are colored in white (vdW representation). The letters close to the Pro rings refers to the respective puckering of its side chain.

Table 2. BE^{*C}, ΔBSSE and ΔE for the various GPP models, e.g. single strand PPII, ss-COL and COL, (as defined in the *Computation Details* section) in kJ·mol^{-1}·triplet^{-1}. London dispersion component of BE^{*C} in parentheses.

Name (XY)	ΔE(PPII)	ΔE(ss-COL)	ΔE(COL)	BE^{*C}	ΔBSSE
DD	0.0	0.0	0.0	82.1 (55.1)	0.0
DU	5.4	1.3	-1.6	84.9 (59.3)	0.1
UD	7.6	6.2	4.8	83.1 (55.2)	0.3
UU	8.5	6.2	5.3	82.7 (56.3)	0.3

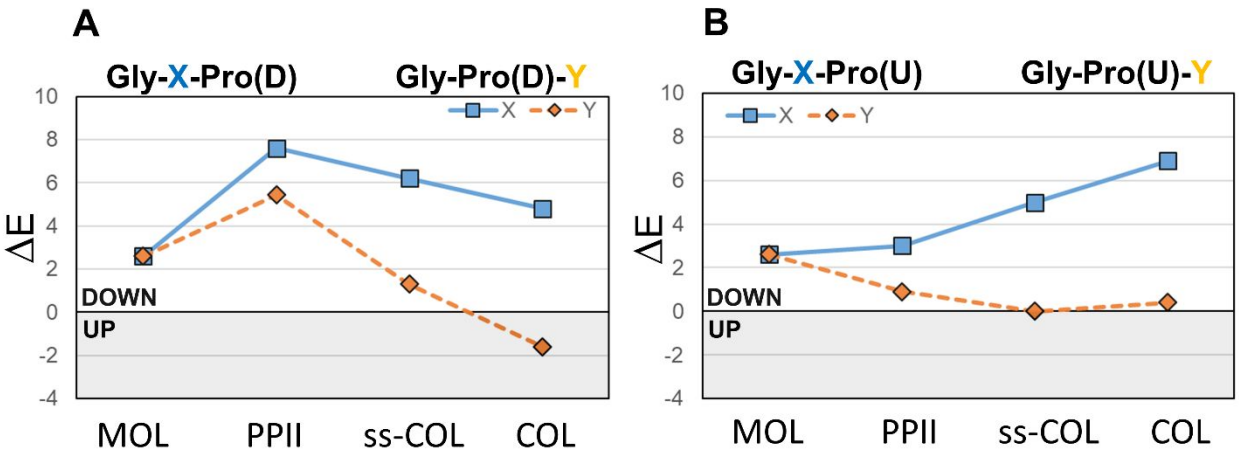


Figure 7. **A-B**: Flipping the Pro puckering from D to U in different protein surrounding for a GPP polymer. In A are reported the cases in which the neighboring Pro is in the D puckering. In B are reported the cases in which the neighboring Pro is in the U puckering. Energy on the ordinate axes in kJ·mol^{-1}·triplet^{-1}. Positive/negative energy value indicate a D/U puckering propensity.

GPO Gas-phase models

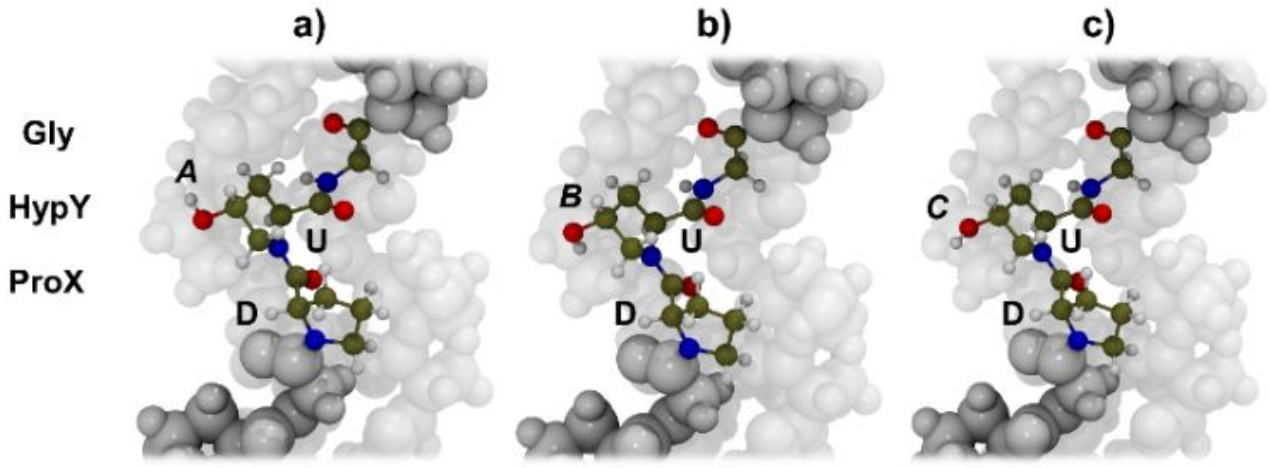
We relaxed the geometry of the ten stable Gly-Pro-Hyp (GPO) unfolded, PPII, and the twelve folded, COL, collagen protein conformers with the B3LYP-D3^{ABC}/VTZP method, arising from the side chain flexibility of Pro and Hyp. The results and the analysis are reported in Figure 8-9 and Table S2-S3. As for the GPP case, the GPO COL inter-strand interaction energy (BE*) as well as the ΔE(PPII), ΔE(COL) and ΔE(ss-COL), are reported in Table 3. The OH group orientation of COL GPO relaxed geometries, see Figure 8 and Table S2, gives conformers whose energy ranking follows the order A < B < C regardless the Pro and Hyp puckering (see Figure 4 for the definition of A, B, C,…). In the most unstable OH orientation, e.g. C, the OH group is sterically hindered, pointing towards the protein core. In the A case there are weak electrostatic contacts with the C=O groups which are either longer or absent for OH in the B conformation, explaining the ranking stability. Therefore, from now on, to simplify the analysis, we will focus only on the COL models within the A oriented OH groups.

For the PPII case, certain OH group orientations give a short intra-strand H-bond whose stabilization dominates the polymer ranking stability (see Table S3 and Figure S1, F OH orientation). We decided to exclude those cases from our analysis because they are expected to become far less relevant in any type of realistic condition (i.e. in water solvation). By looking to the remaining structures, we demonstrate that the OH ranking follows the pattern already seen in the COL case, e.g. A < B < C.

Therefore, as for the COL case, we report in Table 3 and Figure 9 only the PPII polymers with OH group oriented in the A fashion. Among all COL conformers, the DU is the most stable one, in agreement with the GPP case and the experimental evidences (Table 3). In this case, the U→D process in Y position (DU→DD process) is more energetically demanding, e.g. 5.1 kJ·mol^{-1}·triplet^{-1} than in GPP, e.g. 1.6 kJ·mol^{-1}·triplet^{-1}. This was expected for the intrinsic UP puckering propensity of Hyp residues, as we demonstrated for the molecular case (see *Molecular gas-phase models* section). Interestingly, when Hyp is in the U conformation the U→D process (UD→DD process) in X position costs 1.9 kJ·mol^{-1}·triplet^{-1}. This value is reduced with respect to the GPP case, in which it costs 6.9 kJ·mol^{-1}·triplet^{-1}. Conversely, when Hyp is in the D conformation (DD conformer) the D→U process in X position costs as much as in the GPP case, e.g. 4.9 and 4.8 kJ·mol^{-1}·triplet^{-1}. This phenomenon agrees with the experimental evidences gathered from collagen-like peptides. Indeed, in Gly-Pro-Pro (GPP) collagen-like peptides, Pro in the X position prefers a D pucker. Instead, in Gly-Pro-Hyp (GPO) collagen-like peptides (in which Hyp is usually in the U state), the Pro in X position has uniformly distributed pucker among the D/U states (4/3 or 28/23).[55,56] Recently, Chow et al.,[57] by means of solid state nuclear magnetic resonance spectroscopy and classical molecular dynamics calculations related the free ring flipping of Pro in GPO peptides to the presence of Hyp in Y position. Our DFT simulations confirm their findings, moreover indicating that the phenomenon takes places only if Hyp is in the U state. Interestingly for GPO composition, both the ΔE(ss-COL) and ΔE(PPII) values are very similar to those for GPP (see Table 2 and 3). This was not expected, as the puckering propensity of Hyp should reduce the cost of the U puckering in the Y position. Forcing the polymer within the constraints of a PPII helix, induces subtle interactions changing the polymer ranking energy (see the already mentioned inter-strand H-bond). For the PPII geometry, we found out that there might be electrostatic interactions between the OH and the C=O groups of the polymer, leading to these changes in the polymer stability. To check for that, we have changed the substituent of Pro in Y from OH to fluorine. The organic halogen substitution changes the electrostatics contacts within the polymer. At the same time Flp(F) has the same puckering propensity of Hyp (see *Molecular gas-phase models* section). We geometry optimized the GPF PPII polymer resulting in a change in the energy ranking. For GPF, we obtained DD(0.0), DU(0.3), UD(7.5) and UU(5.5 kJ·mol^{-1} triplet^{-1}). The results differ from those of the GPP polymer, in line with the expected U puckering propensity of Flp. As for the BE^{*C} (see Table 3), the values computed for the GPO case are comparable with those of GPP. Interestingly, when Hyp is in the U puckering this leads to a higher inter-strand interaction, mainly due to London dispersion interaction, as for the DU case of GPP COL.

Figure 10 shows the comparison of the relative stability with respect to the most stable conformation for GPP and GPO. The ΔE(COL) numerical data of Table 2 (GPP) and 3 (GPO) have

been used to build up the graphs. For both polymers, the DU state is the most stable one followed by



the same trend DD > UD, with the GPO exhibiting higher instability compared to GPP along that series. The situation changes for the UU case, in which while for GPP the instability is similar to the UD case, for the GPO the value is only 2 kJ mol^{-1} $triplet^{-1}$ higher than the most stable DU conformer. In summary, Figure 10 shows the role of the Pro hydroxylation in Y position to be summarized as:

a) the D puckering in Y position increasing instability compared to the DU state
b) when Hyp is in the U state the D or U puckering in X position becomes close in stability.

Point a) is due to the well-known effect of electronegative substituents on R position of C_γ. Point b) confirms the Chow et al. results:[57] when Hyp is in the U state, the Pro in X position has both D and U states easily accessible. The phenomenon of the "free ring flipping" in GPO peptides has been related to the presence of Hyp in Y position by means of solid state nuclear magnetic resonance spectroscopy and classical molecular dynamics calculations.[57] By means of our simulation this seems to happen only in the case of Hyp in the U state.

We could not carry out the same analysis done for GPP as reported in Figure 7A-B for the GPO composition because, in some cases, there are electrostatic interactions between the Hyp OH group and the C=O groups, possibly due to the absence of solvation. For the Pro side chain analysis, these can be considered spurious interactions. Indeed, they guide the relative stability ranking, thus limiting the possibility of our analysis, as already mentioned before.

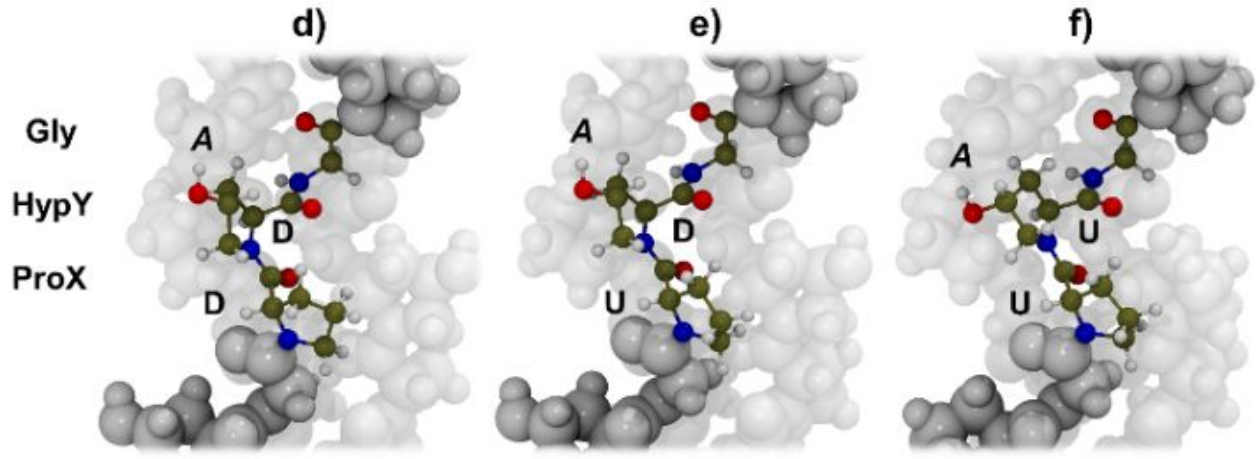


Figure 8. B3LYP-D3^{ABC}/VTZP optimized geometries of the GPO COL models. The asymmetric aminoacid triplet is highlighted in ball-and-stick representation, whilst the other atoms are colored in white (vdW representation). ***a-c)*** Different hydroxyl group orientations for the DU conformation. ***d-f)*** All the possible Pro conformations for the OH in the A orientation. The other cases (B and C orientations of the OH group for DD, UD and UU) are not reported since the differences with the A orientation are negligible, simply exhibiting the OH group rotated in the other directions as in *a-c*.

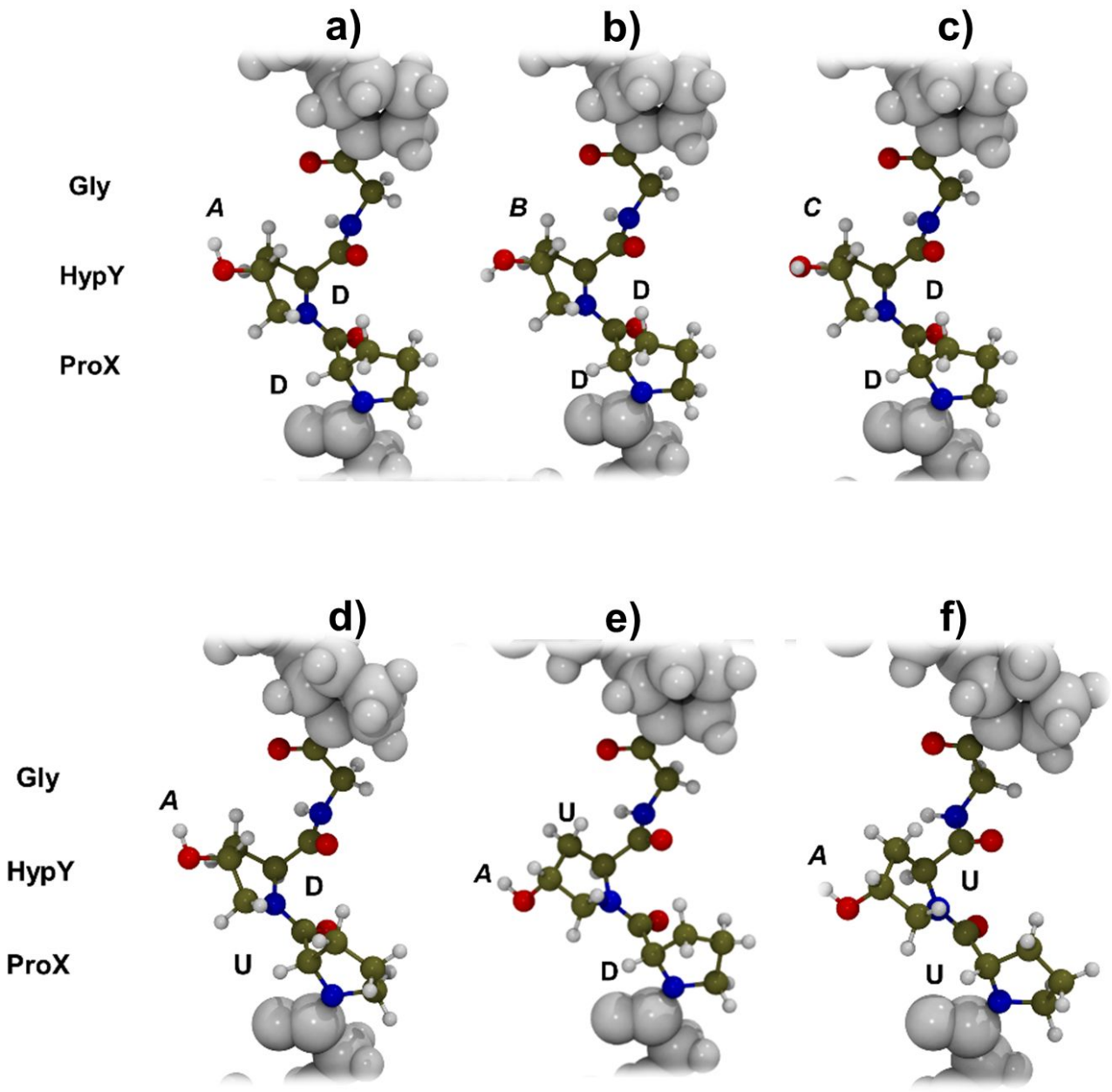


Figure 9. B3LYP-D3^{ABC}/VTZP optimized geometries of the GPO PPII models. The aminoacid asymmetric triplet is highlighted in ball-and-stick representation, while the other atoms are colored in white (vdW representation). *a-c)* Different hydroxyl group orientations for the DD conformation. *d-f)* All the possible Pro/Hyp conformations for the OH in the A orientation. The other cases are reported graphically in Figure S1.

Table 3. BE^{*C}, ΔBSSE and ΔE for the various GPO models, e.g. single strand PPII, ss-COL and COL, (as defined in the *Computation Details* section) in kJ·mol^{-1}·triplet^{-1}. London dispersion component of BE^{*C} in parentheses.

name	ΔE(PPII)	ΔE(ss-COL)	ΔE(COL)	BE^{*C}	ΔBSSE
DD	0.0	0.0	0.0	82.0 (56.1)	0.0
DU	5.9	2.4	-5.1	88.3 (60.0)	1.3
UD	7.3	6.5	4.9	83.3 (56.1)	0.3
UU	13.1	6.5	-3.2	90.6 (59.6)	1.2

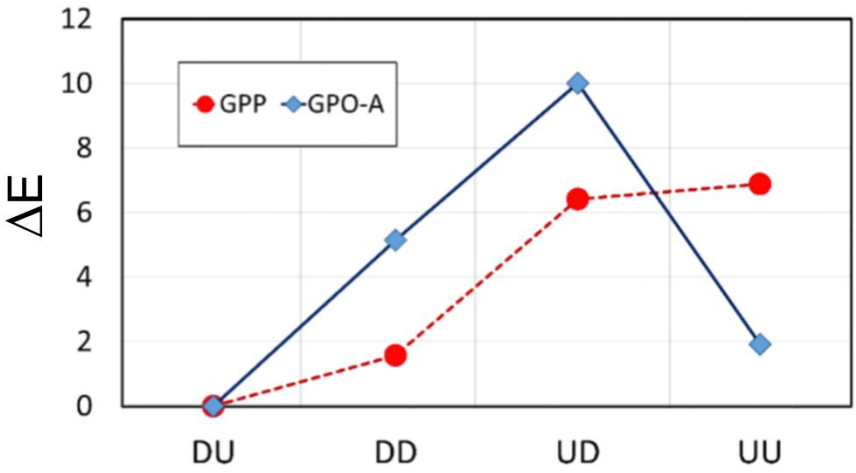


Figure 10. ΔE(COL) relative stability (kJ·mol^{-1}·triplet^{-1}) of GPP and GPO (with A oriented OH group) polymers as a function of the puckering state of the Pro and Hyp aminoacids (see Table 2 and 3 for the numerical values). The labelling indicates the pucker of the Pro in X followed by the one of Pro(GPP)/Hyp(GPO) in Y position.

Conclusions

In this paper we have investigated the effect of Pro conformers on stability and inter-strand energy of hydroxylated and non-hydroxylated collagen triple helices. To clarify the collagen-structure/Pro-side-chain relation, we have also investigated the Pro conformers in different protein organizations. In this way, we have tried to mimic a hypothetical folding process from a random coil to a triple helix. The strength of our work relies on the use of realistic collagen periodic triple helical models coupled with an accurate quantum mechanical approach. The theoretical framework is based on density functional theory through the use of the hybrid B3LYP-D3^{ABC} functional coupled with a large Gaussian basis set. The adopted level of theory together with the use of realistic models allow to accurately account for the small energy differences occurring between conformers dictated by changes in the Pro conformation (D/U states).

We identify three effects which drive Pros to have D/U conformation in X/Y position: i) the Gly presence, which is mandatory within each collagen aminoacidic triplet; ii) the coiled-coiling, in which each single strand has to bend to allow the folding of the protein; iii) the inter-strand interaction arising from electrostatic and dispersive interactions between collagen strands. It seems that the coexistence of these three effects guide the unusual puckering of Pro in collagens, thus giving a sensible physico-chemical base to the propensity based-hypothesis.

Comparing the relative stabilities of GPP and GPO collagens provides insights on the role of Pro hydroxylation on altering the Pro/Hyp conformers accessibility. We can state that hydroxylation of Pro in Y induces: i) the D puckering in Y position to become unstable, due to the well-known effect of electronegative substituents on R position of C_γ; ii) the stabilization of the U state for Pro in X (only when Hyp is in the U state). This confirms the phenomenon of the "free ring flipping" in GPO collagen like peptides.[57] The presence of Hyp in Y position seems to induce equal accessibility of D and U states for Pro in X. By means of our simulation this seems to happen only in the case of Hyp in the U state.

Regarding the binding energy between collagen strands, this depends only slightly on hydroxylation process and side chain conformations on absolute scale. It has values around 83.2 ± 1.2 kJ·mol^{-1}·triplet^{-1} and 86.0 ± 4.1 kJ·mol^{-1}·triplet^{-1} for GPP and GPO, respectively. Interestingly, in all cases, the binding energy is mainly constituted by London dispersion contribution (≈70%). Excluding dispersion from the inter-strand binding energy gives an indication of the strength of the H-bond between the COL strands. The computed H-bond energy is 13.7 ± 0.9 kJ·mol^{-1}, which indicates it as weak H-bond.

The proposed protein-model/level-of-theory combination has demonstrated to be reliable in predicting the correct collagen conformations, in agreement with the large amount of available

experimental data. This approach is easily applicable to homo- and hetero- trimeric collagen with any composition. The reduced cost of the simulations also allows to micro solvate the protein, thus focusing on the protein-solvent interaction. This is a fundamental, as well as delicate step, to shed some light on triple helix stability reasons.[1,18,19] Therefore, we have started studying the water/collagen interaction before and after Pro hydroxylation. Finally, the approach presented here can be used side by side with experimental collagen investigations,[22,58–64] as complementary, reliable, accurate and parameter-free methodology.

Supporting information

CRYSTAL17 Input file for the DU conformation of the GPP Collagen. Table S1: Geometrical parameters and energies of the GPP triple helix (COL) models. Table S4: Geometrical parameters and energies of the GPO triple helix models. Figure S11: Equilibrium geometries of the GPO single strand in the PPII conformation. Table S5. Geometrical parameters and energies of the GPO single strands models.

Acknowledgment

M.C. acknowledges the generous allowance of CINECA computing time from ISCRA B (Project: ISB16; Account ID: MACBONE, Origin ID: HP10BAL7D8), C3S Competence Centre for scientific computing of the University of Torino, B. Civalleri and A. Erba for supporting the usage of CRYSTAL17 program.

References

(1) Bella, J. Collagen Structure: New Tricks from a Very Old Dog. *Biochem. J.* **2016**, *473*, 1001–1025.

(2) Shoulders, M. D.; Raines, R. T. Collagen Structure and Stability. *Annu. Rev. Biochem.* **2009**, *78*, 929–958.

(3) Ramachandran, G. N.; Kartha, G. Structure of Collagen. *Nature* **1954**, *174*, 269–270.

(4) Ramachandran, G. N.; Kartha, G. Structure of Collagen. *Nature* **1955**, *176*, 593–595.

(5) Rich, A.; Crick, F. H. C. The Molecular Structure of Collagen. *J. Mol. Biol.* **1961**, *3*, 483–506.

(6) Cowan, P. M.; Mcgavin, S.; North, A. C. T. The Polypeptide Chain Configuration of Collagen. *Nature* **1955**, *176*, 1062–1064.

(7) Fratzl, P. *Collagen: Structure and Mechanics - an Introduction*; Fratzl, P., Ed.; Springer Berlin / Heidelberg: Postdam, Germany, 2008; pp 1-13.

(8) Crick, F. H. C.; Rick, A. Structure of Polyglycine II. *Nature* **1955**, *176*, 780–781.

(9) Arnott, S.; Dower, S. D. The Structure of Poly-l-Proline II. *Acta Cryst. B* **1968**, *24*, 599–601.

(10) Ramshaw, J. A. M.; Shah, N. K.; Brodsky, B. Gly-X-Y Tripeptide Frequencies in Collagen: A Context for Host-Guest Triple-Helical Peptides. *J. Struct. Biol.* **1998**, *122*, 86–91.

(11) Kapitán, J.; Baumruk, V.; Kopecký, V.; Pohl, R.; Bouř, P. Proline Zwitterion Dynamics in Solution, Glass, and Crystalline State. *J. Am. Chem. Soc.* **2006**, *128*, 13451–13462.

(12) Czinki, E.; Császár, A. G. Conformers of Gaseous Proline. *Chem. - A Eur. J.* **2003**, *9*, 1008–1019.

(13) Vitagliano, L.; Berisio, R.; Mastrangelo, A.; Mazzarella, L.; Zagari, A. Preferred Proline Puckerings in Cis and Trans Peptide Groups: Implications for Collagen Stability. *Protein Sci.* **2010**, *10*, 2627–2632.

(14) Improta, R.; Benzi, C.; Barone, V. Understanding the Role of Stereoelectronic Effects in Determining Collagen Stability. 1. A Quantum Mechanical Study of Proline, Hydroxyproline, and Fluoroproline Dipeptide Analogues in Aqueous Solution. *J. Am. Chem. Soc.* **2001**, *123*, 12568–12577.

(15) Vitagliano, L.; Berisio, R.; Mazzarella, L.; Zagari, A. Structural Bases of Collagen Stabilization Induced by Proline Hydroxylation. *Biopolymers* **2001**, *58*, 459–464.

(16) Motooka, D.; Kawahara, K.; Nakamura, S.; Doi, M.; Nishi, Y.; Nishiuchi, Y.; Kang, Y. K.; Nakazawa, T.; Uchiyama, S.; Yoshida, T.; et al. The Triple Helical Structure and Stability of Collagen Model Peptide with 4(S)-Hydroxyprolyl-Pro-Gly Units. *Biopolymers* **2012**, *98*, 111–121.

(17) Hodges, J. A.; Raines, R. T. Stereoelectronic and Steric Effects in the Collagen Triple Helix: Toward a Code for Strand Association. *J. Am. Chem. Soc.* **2005**, *127*, 15923–15932.

(18) Nishi, Y.; Uchiyama, S.; Doi, M.; Nishiuchi, Y.; Nakazawa, T.; Ohkubo, T.; Kobayashi, Y. Different Effects of 4-Hydroxyproline and 4-Fluoroproline on the Stability of Collagen Triple Helix. *Biochemistry* **2005**, *44*, 6034–6042.

(19) Kawahara, K.; Nishi, Y.; Nakamura, S.; Uchiyama, S.; Nishiuchi, Y.; Nakazawa, T.; Ohkubo, T.; Kobayashi, Y. Effect of Hydration on the Stability of the Collagen-like Triple-Helical Structure of [4(R)-Hydroxyprolyl-4(R)-Hydroxyprolylglycine]$_{10}$. *Biochemistry* **2005**, *44*, 15812–15822.

(20) Domene, C.; Jorgensen, C.; Abbasi, S. A Perspective on Structural and Computational Work on Collagen. *Phys. Chem. Chem. Phys.* **2016**, *18*, 24802–24811.

(21) Zheng, H.; Lu, C.; Lan, J.; Fan, S.; Nanda, V.; Xu, F. How Electrostatic Networks Modulate Specificity and Stability of Collagen. *Proc. Natl. Acad. Sci.* **2018**, *115*, 6207–6212.

(22) Zhang, Y.; Malamakal, R. M.; Chenoweth, D. M. Aza-Glycine Induces Collagen

Hyperstability. *J. Am. Chem. Soc.* **2015**, *137*, 12422–12425.

(23) An, B.; Lin, Y. S.; Brodsky, B. Collagen Interactions: Drug Design and Delivery. *Adv. Drug Deliv. Rev.* **2016**, *97*, 69–84.

(24) Improta, R.; Mele, F.; Crescenzi, O.; Benzi, C.; Barone, V. Understanding the Role of Stereoelectronic Effects in Determining Collagen Stability. 2. A Quantum Mechanical/Molecular Mechanical Study of (Proline-Proline-Glycine)$_n$ Polypeptides. *J. Am. Chem. Soc.* **2002**, *124*, 7857–7865.

(25) Bryan, M. A.; Brauner, J. W.; Anderle, G.; Flach, C. R.; Brodsky, B.; Mendelsohn, R. FTIR Studies of Collagen Model Peptides: Complementary Experimental and Simulation Approaches to Conformation and Unfolding. *J. Am. Chem. Soc.* **2007**, *129*, 7877–7884.

(26) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(27) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(28) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.

(29) Muto, Y. The Force Between Nonpolar Molecules. *Proc. Phys. Math. Soc. Japan* **1944**, *17*, 629–631.

(30) Axilrod, B. M.; Teller, E. Interaction of the van Der Waals Type Between Three Atoms. *J. Chem. Phys.* **1943**, *11*, 299–300.

(31) Schäfer, A.; Huber, C.; Ahlrichs, R. Fully Optimized Contracted Gaussian Basis Sets of Triple Zeta Valence Quality for Atoms Li to Kr. *J. Chem. Phys.* **1994**, *100*, 5829−5835.

(32) Cutini, M.; Corno, M.; Ugliengo, P. Method Dependence of Proline Ring Flexibility in the Poly-L-Proline Type II Polymer. *J. Chem. Theory Comput.* **2017**, *13*, 370−379.

(33) Cutini, M.; Corno, M.; Costa, D.; Ugliengo, P. How Does Collagen Adsorb on Hydroxyapatite? Insights from Ab Initio Simulations on a Polyproline Type II Model. *J. Phys. Chem. C* **2018**, *123*, 7540–7550.

(34) Ruggiero, M. T.; Sibik, J.; Orlando, R.; Zeitler, J. A.; Korter, T. M. Measuring the Elasticity of Poly- l -Proline Helices with Terahertz Spectroscopy. *Angew. Chemie Int. Ed.* **2016**, *55*, 6877–6881.

(35) London, R. E. On the Interpretation of Carbon-13 Spin-Lattice Relaxation Resulting from Ring Puckering in Proline. *J. Am. Chem. Soc.* **1978**, *100*, 2678–2685.

(36) Sone, M.; Yoshimizua, H.; Kurosu, H.; Ando, I. Side-Chain Conformation of Poly (L-Proline) Form II in the Crystalline State as Studied by High-Resolution Solid-State ^{13}C NMR

Spectroscopy. *J. Molec. Struct.* **1994**, *17*, 111–118.

(37) Kapitán, J.; Baumruk, V.; Bouř, P. Demonstration of the Ring Conformation in Polyproline by the Raman Optical Activity. *J. Am. Chem. Soc.* **2006**, *128*, 2438–2443.

(38) Mádi, Z. L.; Griesinger, C.; Ernst, R. R. Conformational Dynamics of Proline Residues in Antamanide. J Coupling Analysis of Strongly Coupled Spin Systems Based on ECOSY Spectra. *J. Am. Chem. Soc.* **1990**, *112*, 2908–2914.

(39) Sarkar, S. K.; Young, P. E.; Torchia, D. A. Ring Dynamics of DL-Proline and DL-Proline Hydrochloride in the Solid State: A Deuterium Nuclear Magnetic Resonance Study. *J. Am. Chem. Soc.* **1986**, *108*, 6459–6464.

(40) Dovesi, R.; Erba, A.; Orlando, R.; Zicovich-Wilson, C. M.; Civalleri, B.; Maschio, L.; Rérat, M.; Casassa, S.; Baima, J.; Salustro, S. *et al* Quantum-Mechanical Condensed Matter Simulations with CRYSTAL. *WIREs Comput Mol Sci.* **2018**, *8:e1360*, 1–36.

(41) Schäfer, A.; Horn, H.; Ahlrichs, R. Fully Optimized Contracted Gaussian Basis Sets for Atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571–2577.

(42) Corno, M.; Busco, C.; Civalleri, B.; Ugliengo, P. Periodic Ab Initio Study of Structural and Vibrational Features of Hexagonal Hydroxyapatite $Ca_{10}(PO_4)_6(OH)_2$. *Phys. Chem. Chem. Phys.* **2006**, *8*, 2464–2472.

(43) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104–154119.

(44) Broyden, C. G. The Convergence of a Class of Double-Rank Minimization Algorithms 1. General Considerations. *IMA J. Appl. Math.* **1970**, *6*, 76–90.

(45) Fletcher, R. A. New Approach to Variable Metric Algorithms. *Comput. J.* **1970**, *13*, 317–322.

(46) Shanno, D. F.; Kettler, P. C. Optimal Conditioning of Quasi-Newton Methods. *Math. Comput.* **1970**, *24*, 657–664.

(47) Monkhorst, H. J.; Pack, J. D. Special Points for Brillouin-Zone Integration. *Phys. Rev. B.* **1976**, *8*, 5188–5192.

(48) Dovesi, R.; Saunders, V. R.; Roetti, C.; Orlando, R.; Zicovich-Wilson, C. M.; Pascale, F.; Civalleri, B.; Doll, K.; Harrison, N. M.; Bush, I. J.; et al. *CRYSTAL14, User's Manual*; Università di Torino: Torino, Italy, 2014.

(49) Prencipe, M.; Pascale, F.; Zicovich-Wilson, C. M.; Saunders, V. R.; Orlando, R.; Dovesi, R. The Vibrational Spectra of Calcite ($CaCO_3$): An Ab Inito Quantum-Mechanical Calculation. *Phys. Chem. Miner.* **2004**, *31*, 1–6.

(50) Ugliengo, P.; Viterbo, D.; Chiari, G. MOLDRAW: Molecular Graphics on a Personal

Computer. *Z. Krist.* **1993**, *207*, 9–23.

(51) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(52) Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A. Crystal Structure of the Collagen Triple Helix Model $[(Pro\text{-}Pro\text{-}Gly)_{10}]_3$. *Protein Sci.* **2002**, *11*, 262–270.

(53) Hongo, C.; Nagarajan, V.; Noguchi, K.; Kamitori, S.; Okuyama, K.; Tanaka, Y.; Nishino, N. Average Crystal Structure of $(Pro\text{-}Pro\text{-}Gly)_9$ at 1.0 Å Resolution. *Polym. J.* **2001**, *33*, 812–818.

(54) Okuyama, K.; Nagarajan, V.; Kamitori, S. 7/2-Helical Model for Collagen - Evidence from Model Peptides. *Proc. Indian Acad. Sci. - Chem. Sci.* **1999**, *111*, 19–34.

(55) Okuyama, K.; Miyama, K.; Mizuno, K.; Bächinger, H. P. Crystal Structure of (Gly-Pro-Hyp)9: Implications for the Collagen Molecular Model. *Biopolymers* **2012**, *97*, 607–616.

(56) Okuyama, K.; Hongo, C.; Fukushima, R.; Wu, G.; Narita, H.; Noguchi, K.; Tanaka, Y.; Nishino, N. Crystal Structures of Collagen Model Peptides with Pro-Hyp-Gly Repeating Sequence at 1.26 Å Resolution: Implications for Proline Ring Puckering. *Biopolym. - Pept. Sci. Sect.* **2004**, *76*, 367–377.

(57) Chow, W. Y.; Bihan, D.; Forman, C. J.; Slatter, D. A.; Reid, D. G.; Wales, D. J.; Farndale, R. W.; Duer, M. J. Hydroxyproline Ring Pucker Causes Frustration of Helix Parameters in the Collagen Triple Helix. *Sci. Rep.* **2015**, *5* (February), 12556.

(58) Egli, J.; Siebler, C.; Maryasin, B.; Erdmann, R. S.; Bergande, C.; Ochsenfeld, C.; Wennemers, H. PH-Responsive Aminoproline-Containing Collagen Triple Helices. *Chem. - A Eur. J.* **2017**, *23*, 7938–7944.

(59) Aronoff, M. R.; Egli, J.; Menichelli, M.; Wennemers, H. γ-Azaproline Confers PH Responsiveness and Functionalizability on Collagen Triple Helices. *Angew. Chemie - Int. Ed.* **2019**, *58*, 3143–3146.

(60) Egli, J.; Siebler, C.; Köhler, M.; Zenobi, R.; Wennemers, H. Hydrophobic Moieties Bestow Fast-Folding and Hyperstability on Collagen Triple Helices. *J. Am. Chem. Soc.* **2019**, *141*, 5607–5611.

(61) Kasznel, A. J.; Zhang, Y.; Hai, Y.; Chenoweth, D. M. Structural Basis for Aza-Glycine Stabilization of Collagen. *J. Am. Chem. Soc.* **2017**, *139*, 9427–9430.

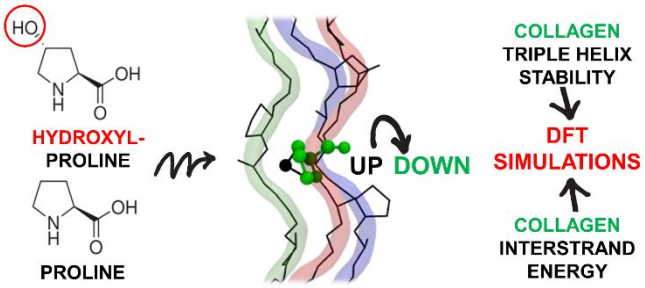
(62) Hentzen, N. B.; Smeenk, L. E. J.; Witek, J.; Riniker, S.; Wennemers, H. Cross-Linked Collagen Triple Helices by Oxime Ligation. *J. Am. Chem. Soc.* **2017**, *139*, 12815–12820.

(63) Tanrikulu, I. C.; Raines, R. T. Optimal Interstrand Bridges for Collagen-like Biomaterials. *J. Am. Chem. Soc* **2014**, *136*, 13490−13493.

(64) Testa, A.; Lucas, X.; Castro, G. V.; Chan, K. H.; Wright, J. E.; Runcie, A. C.; Gadd, M. S.;

Harrison, W. T. A.; Ko, E. J.; Fletcher, D.; et al. 3-Fluoro-4-Hydroxyprolines: Synthesis, Conformational Analysis, and Stereoselective Recognition by the VHL E3 Ubiquitin Ligase for Targeted Protein Degradation. *J. Am. Chem. Soc.* **2018**, *140*, 9299–9313.

TOC Graphic


HO
OH
N
H
O
HYDROXYL-
PROLINE
OH
N
H
O
PROLINE
UP
DOWN
COLLAGEN
TRIPLE HELIX
STABILITY
DFT
SIMULATIONS
COLLAGEN
INTERSTRAND
ENERGY